Mathematical Modeling of Behavior (2024-2025)

Michel Bierlaire.

TAs: Cloe Cortes, Pavel Ilinov, Evangelos Paschalidis, Anne-Valérie Preto, Negar Rezvany

Mid-term wrap-up (correction) — November 12, 2024

Question 1: Specification testing and forecasting

Interpretation of the specification

1. They are the alternative specific constants. The parameter $\mathrm{ASC}_{\mathrm{car}}$ captures the mean of the difference of the error term of the car alternative and the error term of the Swissmetro alternative:

   $$\mathrm{ASC}_{\mathrm{car}} = \mathrm{E}[\varepsilon_{\mathrm{car},n} - \varepsilon_{\mathrm{SM},n}] = \mathrm{E}[\varepsilon_{\mathrm{car},n}] - \mathrm{E}[\varepsilon_{\mathrm{SM},n}].$$

   The parameter $\mathrm{ASC}_{\mathrm{rail}}$ is similarly defined:

   $$\mathrm{ASC}_{\mathrm{rail}} = \mathrm{E}[\varepsilon_{\mathrm{rail},n} - \varepsilon_{\mathrm{SM},n}] = \mathrm{E}[\varepsilon_{\mathrm{rail},n}] - \mathrm{E}[\varepsilon_{\mathrm{SM},n}].$$

2. The utilities are not observed. Only the choices are observed. Therefore, it is not possible to identify the three constants. Only their difference can be identified. Therefore, one of them must be normalized to 0. In this model, it has been decided to normalize the constant of the Swissmetro alternative to 0.

3. The two parameters have a negative sign. It means that, when the corresponding variable increases, the utility function decreases, and so does the choice probability.

4. The terms $\beta_{\mathrm{f,car}}\mathrm{female}_n$ and $\beta_{\mathrm{f,rail}}\mathrm{female}_n$ capture an interaction between the constants and the gender. It is meant to relax the hypothesis that the constants are the same for all members of the population. In this model, the constants are different depending on gender. As they interact with the constant, they must be normalized in the same way as the constant, so that no such term appears in the Swissmetro alternative.

5. If n is female, then $\mathrm{female}_n = 1$, and the alternative specific constant for car is

   $$\mathrm{ASC}_{\mathrm{car}} + \beta_{\mathrm{f,car}}\mathrm{female}_n = \mathrm{ASC}_{\mathrm{car}} + \beta_{\mathrm{f,car}} = -0.461 + 0.309 = -0.152.$$

   The alternative specific constant for rail is

   $$\mathrm{ASC}_{\mathrm{rail}} + \beta_{\mathrm{f,rail}}\mathrm{female}_n = \mathrm{ASC}_{\mathrm{rail}} + \beta_{\mathrm{f,rail}} = 0.0906 - 1.23 = -1.14.$$

6. If n is male, then $\text{female}_n = 0$, and the alternative specific constant for car is

$$\text{ASC}_{\text{car}} + \beta_{\text{f,car}}\text{female}_n = \text{ASC}_{\text{car}} = -0.461.$$

The alternative specific constant for rail is

$$\text{ASC}_{\text{rail}} + \beta_{\text{f,rail}}\text{female}_n = \text{ASC}_{\text{rail}} = 0.0906.$$

7. The specification of the model is

$$\begin{aligned} V_{\text{car},n} &= \text{ASC}_{\text{car}} + \beta_c\text{cost}_{\text{car},n} + \beta_t\text{time}_{\text{car},n} + \beta_{\text{m,car}}\text{male}_n, \\ V_{\text{rail},n} &= \text{ASC}_{\text{rail}} + \beta_c\text{cost}_{\text{rail},n} + \beta_t\text{time}_{\text{rail},n} + \beta_{\text{m,rail}}\text{male}_n, \\ V_{\text{SM},n} &= \beta_c\text{cost}_{\text{SM},n} + \beta_t\text{time}_{\text{SM},n}. \end{aligned}$$

The value of the coefficients of the cost and time variables are exactly the same as for model M_1. The values of the constants, $\beta_{\text{m,car}}$ and $\beta_{\text{m,rail}}$ are calculated using the values of the constants for each segment calculated above.

If n is female, then $\text{male}_n = 0$, and the alternative specific constant for car is

$$\text{ASC}_{\text{car}} + \beta_{\text{m,car}}\text{male}_n = \text{ASC}_{\text{car}} = -0.152.$$

The alternative specific constant for rail is

$$\text{ASC}_{\text{rail}} + \beta_{\text{m,rail}}\text{male}_n = \text{ASC}_{\text{rail}} = -1.14.$$

If n is male, then $\text{male}_n = 1$, and the alternative specific constant for car is

$$\text{ASC}_{\text{car}} + \beta_{\text{m,car}}\text{male}_n = \text{ASC}_{\text{car}} + \beta_{\text{m,car}} = -0.461.$$

Therefore,

$$\beta_{\text{m,car}} = -0.461 - (-0.152) = -0.309.$$

The alternative specific constant for rail is

$$\text{ASC}_{\text{rail}} + \beta_{\text{m,rail}}\text{male}_n = \text{ASC}_{\text{rail}} + \beta_{\text{m,rail}} = 0.0906.$$

Therefore,

$$\beta_{\text{m,rail}} = 1.23.$$

Therefore, we obtain the results in Table 1 on the following page

ASC_{car}	-0.152
ASC_{rail}	-1.14
β_c	-0.0108
β_t	-0.0125
$\beta_{m,car}$	-0.309
$\beta_{m,rail}$	1.23

Table 1: Parameter estimates for M_2

8. The specification of the model is

$$\begin{aligned}
V_{car,n} &= \beta_{f,car}\text{female}_n + \beta_{m,car}(1-\text{female}_n) + \beta_c\text{cost}_{car,n} + \beta_t\text{time}_{car,n},\\
V_{rail,n} &= \beta_{f,rail}\text{female}_n + \beta_{m,rail}(1-\text{female}_n) + \beta_c\text{cost}_{rail,n} + \beta_t\text{time}_{rail,n},\\
V_{SM,n} &= \beta_c\text{cost}_{SM,n} + \beta_t\text{time}_{SM,n}.
\end{aligned}$$

The value of $\beta_{f,car}$, $\beta_{m,car}$, $\beta_{f,rail}$, $\beta_{m,rail}$ have been calculated above:

$$\begin{aligned}
\beta_{f,car} &= -0.152,\\
\beta_{m,car} &= -0.461,\\
\beta_{f,rail} &= -1.14,\\
\beta_{m,rail} &= 0.0906.
\end{aligned}$$

The value of all the other parameters are the same as for model M_1. Note that the model can also be written as

$$\begin{aligned}
V_{car,n} &= \beta_{m,car} + (\beta_{f,car} - \beta_{m,car})\text{female}_n + \beta_c\text{cost}_{car,n} + \beta_t\text{time}_{car,n},\\
V_{rail,n} &= \beta_{m,rail} + (\beta_{f,rail} - \beta_{m,rail})\text{female}_n + \beta_c\text{cost}_{rail,n} + \beta_t\text{time}_{rail,n},\\
V_{SM,n} &= \beta_c\text{cost}_{SM,n} + \beta_t\text{time}_{SM,n}.
\end{aligned}$$

We see that the parameters $\beta_{m,car}$ and $\beta_{m,rail}$ in M_3 correspond to ASC_{car} and ASC_{rail} in M_1. The parameters $\beta_{f,car}$ and $\beta_{f,rail}$ in M_1 correspond to $(\beta_{f,car} - \beta_{m,car})$ and $(\beta_{f,rail} - \beta_{m,rail})$ in M_3.

Testing

1. We say that a parameter is significant when we can reject the null hypothesis that its true value is equal to zero, which would mean that the corresponding variable does not play any role in the model. The corresponding test is a t-test. *Note: the concept of "being significant" is associated with a confidence level. For instance, at a 5% confidence level, a coefficient is not significant when the corresponding* t*-ratio in absolute value is lower than 1.96 (*$|\text{t-ratio}| < 1.96$*) and, equivalently, when the corresponding* p*-value is larger than 0.05 (*$|\text{p-value}| > 0.05$*). Note that, in specification testing, the concept of "being significant" applies only when the null hypothesis to be tested makes sense. Also, it is usually better to apply higher confidence levels than in classical hypothesis testing, as "type II" errors (that is, specification errors) are more damaging than "type I" errors (that is, presence of an irrelevant variable).*

   In the case of model M_4, there are two coefficients with a p-value larger than 0.05: ASC_{car} and $\beta_{t,car}$. However, we cannot declare them "not significant", as it does not make sense to test the null hypothesis that they are equal to zero:

   - The constant is capturing the difference of the mean of two error terms. A true value of zero would mean that the means are equal, which is not a relevant hypothesis to test in this context. The fact that this difference is numerically close to zero has no concrete behavioral meaning.
   - The hypothesis that the coefficient $\beta_{t,car}$ is equal to zero is equivalent to assuming that the travel time variable does not play any role in the model. This does not make sense, as travel time is a key explanatory variable in transportation mode choice models. The low t-test (or the high p-value) is a consequence of a high standard error, probably due to a lack of variability in the data. In this case, it is a sign that more (or better) data is needed.

2. The likelihood ratio test can be used because model M_1 is a restricted version of model M_4. The null hypothesis is "the restricted model M_1 is the true model". Five linear restrictions must be applied to M_4 in order to obtain M_1:

$$\begin{aligned} \beta_{\text{GA,SM}} = \beta_{\text{GA,rail}} &= 0, \\ \beta_{\text{t,car}} = \beta_{\text{t,rail}} &= \beta_{\text{t,SM}}, \\ \lambda &= 1. \end{aligned}$$

3. The statistic for the likelihood ratio test is

$$-2(\mathcal{L}^R - \mathcal{L}^U) = -2(\mathcal{L}^1 - \mathcal{L}^4) = -2(-5187.983 + 4936.917) = 502.132.$$

Because there are five linear restrictions, we need to compare this value with the 95% quantile of the χ^2 distribution with five degrees of freedom, which is 11.07. Indeed, if X follows a χ^2 distribution with five degrees of freedom, we have

$$\Pr(X \leq 11.07) = 0.95,$$

or, equivalently

$$\Pr(X \geq 11.07) = 0.05.$$

As the value of the test, 502.132, exceeds by far the threshold, we can safely reject the null hypothesis at the 5% level of significance, and M_4 is preferred to M_1.

4. The fact that the marginal effect of travel cost in the utility varies with travel cost means that the utility function is non linear in travel cost. Therefore, we can replace the cost variable by any non linear transformation of it. Note that the corresponding model is linear-in-parameters, and the new variable is simply associated with a coefficient, like in the original specification.

As an example for M_5, we consider a logarithmic transformation:

$$\begin{aligned} V_{\text{car},n} &= \cdots + \beta'_{\text{cost}} \ln(\text{cost}_{\text{car},n}) + \dots \\ V_{\text{rail},n} &= \cdots + \beta'_{\text{cost}} \ln(\text{cost}_{\text{rail},n}) + \dots \\ V_{\text{SM},n} &= \cdots + \beta'_{\text{cost}} \ln(\text{cost}_{\text{SM},n}) + \dots \end{aligned}$$

5. Comparing M_4 and M_5 cannot be done using a likelihood ratio test, as no model is a restricted version of the other. We refer to this context as testing "non nested hypotheses". Two tests can be performed:

   (a) The Cox test consists in estimating a composite model such that both M_4 and M_5 are restricted versions of this model. Therefore, likelihood ratio tests can be used to test M_4 and M_5 against the composite model.

   (b) To test the hypothesis that the true model is M_4, say, the Davidson-McKinnon J-test consists in first estimating the parameters of M_5, and in including the estimated value of each utility function as an explanatory variable in the specification of the corresponding utility function of M_4. If the hypothesis is true, this additional variable should not play any role, and its coefficient should not be significantly different from zero. A symmetric procedure is applied to test the hypothesis that M_5 is the true model.

   For each of these tests, there are three possible outcomes:

   - one of the two models is rejected, and we keep the other one,
   - both models are rejected, and we investigate better models,
   - no model is rejected, and we use an adjusted likelihood ratio index to select among them ($\bar{\rho}^2$, AIC or BIC).

6. The value of time associated with each alternative for model M_5 is calculated using the definition

$$\text{VOT}_{in} = \frac{\partial V_{in}/\partial \text{time}_{in}}{\partial V_{in}/\partial \text{cost}_{in}}.$$

In the case of the logarithm specification:

$$\text{VOT}_{\text{rail},n}(\text{cost}_{\text{rail},n}) = \frac{\beta_{\text{time,rail}}}{\frac{\beta'_{\text{cost}}}{\text{cost}_{\text{rail},n}}} = \frac{\beta_{\text{time,rail}} \cdot \text{cost}_{\text{rail},n}}{\beta'_{\text{cost}}},$$

$$\text{VOT}_{\text{SM},n}(\text{cost}_{\text{SM},n}) = \frac{\beta_{\text{time,SM}}}{\frac{\beta'_{\text{cost}}}{\text{cost}_{\text{SM},n}}} = \frac{\beta_{\text{time,SM}} \cdot \text{cost}_{\text{SM},n}}{\beta'_{\text{cost}}},$$

$$\text{VOT}_{\text{car},n}(\text{cost}_{\text{car},n}, \text{time}_{\text{car},n}) = \frac{\beta_{\text{time,car}}\text{time}^{\lambda-1}_{\text{car},n}}{\frac{\beta'_{\text{cost}}}{\text{cost}_{\text{car},n}}} = \frac{\beta_{\text{time,car}}\text{time}^{\lambda-1}_{\text{car},n} \cdot \text{cost}_{\text{car},n}}{\beta'_{\text{cost}}}.$$

The value of time VOT_{in} represents the price (in CHF in this case) that individual n is willing to pay to save one unit (minute in this case) of travel time with alternative i.